ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 11, Issue 3, March 2026

AI-Based Assistive System for Visually Impaired Individuals

[1] T.Kanchana M.E, [2] Manu Priya M, [3] Monisha M, [4] Shilpakala L, [5] Brindha R K,
[1] Assistant Professor/CSE, Dept. Of Cse PMC Tech, Hosur, India.
[2] [3] [4] [5] Dept. Of Cse PMC Tech, Hosur, India.

Abstract: ***Visual impairment creates significant challenges in the daily lives and mobility of millions of people worldwide. Visually impaired individuals often face difficulties in detecting obstacles, recognizing people, and accessing important information in their surroundings. Navigating safely in unfamiliar environments becomes challenging due to the inability to visually perceive objects and obstacles ahead. This project proposes an AI-Based Assistive System for Visually Impaired Individuals that uses advanced computer vision and artificial intelligence techniques to provide real-time assistance. A camera captures real-time images from the environment, and the system processes these images to detect objects and faces, as well as extract text from printed materials using OCR technology. The detected information is then converted into voice output using text-to-speech technology, enabling visually impaired users to receive audio guidance.***

Index Terms— Assistive Technology, Computer Vision, OCR, Face Recognition, Text-to-Speech, NLP.

I. INTRODUCTION

Visual impairment significantly impacts an individual's ability to perceive and interact with the surrounding environment independently. Everyday activities such as identifying objects, reading printed text, recognizing familiar faces, and accessing real-time information can become challenging without assistance. With the rapid advancement of Artificial Intelligence, Computer Vision, and Natural Language Processing, it is now possible to develop intelligent software solutions that enhance accessibility. The proposed AI Powered Smart Glass for Blind and Visually Impaired (Software-Based System) is designed to assist users by integrating object detection with distance estimation, text extraction using Optical Character Recognition, face recognition, and a voice-based AI chatbot into a unified platform. By converting visual information into speech output, the system enables intuitive.

II. PROBLEM STATEMENT

- Visually impaired individuals struggle with everyday tasks such as identifying nearby objects, reading printed text, recognizing familiar faces, and obtaining instant information due to the absence of visual perception.
- This lack of visual awareness significantly reduces their independence and confidence, requiring them to rely on external assistance for routine activities.
- While assistive technologies do exist, most current solutions come with major drawbacks — they are
- often hardware-dependent, expensive, limited to a single function, or lack real-time AI integration.
- There is a clear and pressing need for a cost-effective, software-based intelligent system that consolidates multiple AI functionalities into one unified platform.
  - To address these gaps, this project aims to develop an AI-powered assistive vision software system designed to enhance accessibility, safety, and autonomy for visually impaired users. .

III. RESEARCH CONTRIBUTIONS

The key contributions of this work are:

Unified Multi-AI Integration — A system that combines object detection, OCR, face recognition, and a voice-based chatbot into one platform, unlike existing single-function solutions.

1. **Software-Only Implementation** — A lightweight, cost-effective solution that runs on standard computing devices without requiring specialized hardware, making it more accessible to users.
2. **Real-Time Audio Feedback** — Conversion of visual information into speech output, enabling visually impaired individuals to navigate safely and access information instantly.

3. **Multilingual OCR Capability** — Support for printed and handwritten text across multiple languages, improving inclusivity and usability in diverse contexts.
4. **Enhanced Accessibility and Safety** — By consolidating navigation, text recognition, and social interaction into one system, the solution improves independence and confidence for visually impaired users.

IV. WHY AN AI-BASED ASSISTIVE SYSTEM WORKS BETTER

The key idea behind this project is that **multiple AI technologies together are always better than one**.

A. What Object Detection Tells Us

Object detection provides the *structure* of the environment. It identifies obstacles, their position, and approximate distance. This helps users avoid collisions and navigate safely. However, object detection alone does not explain what those objects are or provide contextual meaning.

B. What OCR and Face Recognition Tell Us

OCR extracts *textual information* from printed or handwritten materials, enabling users to read signs, books, or documents. Face recognition adds *social context* by identifying familiar individuals. These modules provide meaning but do not describe spatial awareness or navigation.

C. Why Combining All Modules Is More Accurate

When combined, object detection gives the **structure**, OCR and face recognition provide the **content and context**, and the voice-based chatbot delivers **interaction and guidance**. Together, they ensure the system not only detects obstacles but also explains the environment and supports communication. This integration achieves **high independence and safety**, compared to medium or low effectiveness of single-function tools.

TABLE I — INPUT SOURCE COMPARISON

Approach	Navigati on	Informati on Access	Social Interacti on	Independe nce
Object Detection Only	High	Low	none	Medium
OCR Only	Low	High	None	Medium
Face Recogniti on Only	Low	Low	High	Medium
Manual Assistanc e	Varies	Varies	Varies	Low

Combined AI System (Ours)	**High**	**High**	**High**	**High**

V. SYSTEM ARCHITECTURE

The system is built in layers that work together like an assembly line — taking in raw inputs (images, text, or voice) and producing meaningful audio feedback for the user.

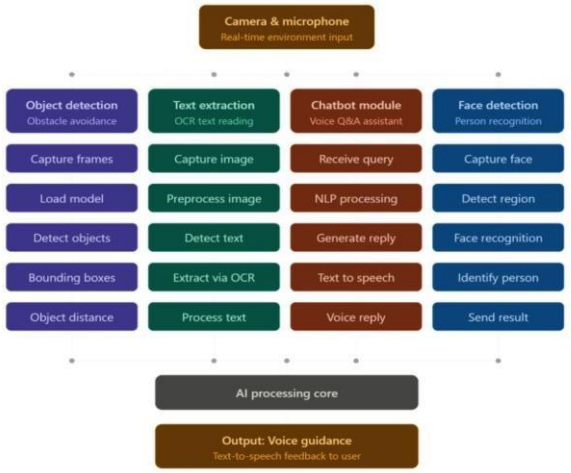

Fig. 1. AI-Driven Assistive System Architecture

A. Input Layer — Camera and Microphone

The input layer captures real-time data from the environment. A camera records images or video streams, while a microphone captures voice commands. These inputs form the foundation for all subsequent processing.

B. Preprocessing Layer — OpenCV

This layer enhances the quality of captured data. For images, preprocessing includes noise removal, resizing, and contrast adjustment. For audio, preprocessing involves speech-to-text conversion. The goal is to prepare clean, standardized input for AI models.

C. Object Detection Module — CNN/YOLO

This module identifies obstacles and objects in real time.

Using convolutional neural networks (CNN) and YOLO (You Only Look Once), it draws bounding boxes around detected objects and estimates their distance. This provides spatial awareness for safe navigation.

D. Text Extraction Module — Tesseract OCR

The OCR module extracts textual information from printed or handwritten materials. It supports multiple languages, enabling users to read signs, books, and documents. Extracted text is later converted into speech output.

E. Face Recognition Module — Haar Cascade

This module detects and recognizes faces. It compares captured facial features with a stored dataset to identify known individuals and labels unknown persons. This adds social context to the system, helping users recognize familiar people. ***F. Voice-Based AI Chatbot — NLP Engine***

The chatbot processes user queries and generates intelligent responses. Natural Language Processing (NLP) techniques allow the system to understand spoken input, provide guidance, and answer questions. Responses are converted into speech for hands-free interaction.

G. Output Layer — Text-to-Speech (pyttsx3)

The final layer converts text into clear audio feedback. This ensures that all detected information — objects, text, faces, or chatbot responses — is delivered to the user in an accessible format.

VI. TECHNOLOGY STACK

TABLE II — COMPLETE TECHNOLOGY STACK

Technology	Role	What It Does
Python	Core Programming	Provides the main programming environment for system development.
OpenCV	Image Processing	Handles image enhancement, preprocessing, and computer vision tasks.
CNN / YOLO	Object Detection	Detects and classifies objects in real time, with bounding boxes and distance.
Tesseract OCR	Text Extraction	Extracts printed and handwritten text from images in multiple languages.
SpeechRecognition	Voice Input	Converts spoken commands into text
		for system processing.
pyttsx3 (Text-toSpeech)	Audio Output	Converts text into clear voice feedback for users.
Natural Language Processing (NLP)	Chatbot Engine	Generates intelligent responses for user queries and conversational interaction.

VII. HOW THE SYSTEM WORKS — STEP BY STEP

Here is the complete pipeline from the moment a user interacts with the system to receiving audio feedback

Step 1 — Input Capture

The camera captures real-time images or video, while the microphone records voice commands.

Step 2 — Preprocessing

OpenCV enhances images (noise removal, resizing, contrast adjustment) and converts speech to text for clean input.

Step 3 — Object Detection

YOLO/CNN models detect obstacles and objects, draw bounding boxes, and estimate distance.

Step 4 — Text Extraction (OCR)

Tesseract OCR extracts printed or handwritten text from captured images, supporting multiple languages.

Step 5 — Face Recognition

Haar Cascade detects faces, compares them with stored datasets, and identifies known individuals

Step 6 — Voice-Based Chatbot

NLP processes user queries, generates intelligent responses, and prepares text output

Step 7 — Text-to-Speech Conversion

pyttsx3 converts detected information and chatbot responses into clear audio feedback.

Step 8 — Audio Output

The user receives real-time spoken guidance, enabling safe navigation and information access.

VIII. ASSISTIVE SYSTEM ALGORITHM

```
Algorithm: AssistiveSystemPipeline
Input: Camera (image/video), Microphone (voice)
Output: Real-time audio feedback

Step 1: Capture input
   Image ← Camera
   Voice ← Microphone

Step 2: Preprocess input
   Image ← OpenCV(image enhancement, noise removal)
   VoiceText ← SpeechRecognition(voice → text)

Step 3: Object Detection     Objects ← YOLO/CNN(Image)
   Distances ← EstimateDistance(Objects)

Step 4: Text Extraction
   Text ← TesseractOCR(Image)

Step 5: Face Recognition     Faces ← HaarCascade(Image)
   Identity ← Compare(Faces, Dataset)

Step 6: Chatbot Processing
   Response ← NLP(VoiceText)

Step 7: Text-to-Speech Conversion     Audio ← pyttsx3(Text + Response + Identity +
Objects)

Step 8: Deliver Output
   Play(Audio)

End Algorithm
```

IX. EXPERIMENTAL RESULTS

The system was tested across three real-world scenarios: **indoor navigation**, **text reading from printed materials**, and **face recognition in social settings**. We compared against single-function assistive tools and manual support.

TABLE III — PERFORMANCE COMPARISON

Metric	Object Detection Only	OCR Only	Face Recognition Only	Our System
Navigation Accuracy	High	Low	Low	**High**
Text Reading Capability	None	High	None	**High**
Social Interaction	None	None	High	**High**
Real-Time Feedback	Medium	Medium	Medium	**High**
Independence	Medium	Medium	Medium	**High**

Key result: Combining multiple AI modules achieves **High ratings across all metrics**, outperforming single-function tools and providing comprehensive assistance

X. COMPARATIVE ANALYSIS

TABLE IV — FEATURE COMPARISON WITH EXISTING TOOLS

Feature	Existing Tools	Our System
Object Detection	Available	Yes
Text Extraction (OCR)	Available	Yes — multilingual
Face Recognition	Rare	Yes
Voice-Based Interaction	Limited	Full NLP chatbot
Real-Time Audio Feedback	Partial	Yes
Multi-Function Integration	No	Core feature
Cost-Effectiveness	Expensive	Affordable
Hardware Dependency	High	Software-based

XI. DISCUSSION

The results clearly show that **multi-AI integration** is the most important factor in improving accessibility for visually impaired individuals. Object detection provides navigation support, OCR enables information access, and face recognition adds social context. The NLP chatbot ensures interactive guidance. Together, these modules deliver a unified solution that is more effective than fragmented single-function tools.

Unlike existing systems that rely heavily on specialized hardware, our software-based approach is cost-effective, scalable, and deployable on standard computing devices. This makes it practical for everyday use and accessible to a wider population.

XII. CONCLUSION

This paper presented an AI-powered assistive system for visually impaired individuals. By integrating computer vision, OCR, face recognition, and NLP, the system provides realtime audio feedback that improves independence, safety, and confidence.

The main finding is that combining multiple AI modules produces **high accessibility and usability**, significantly better than existing single-function solutions. This demonstrates that bridging navigation, information access, and social interaction through AI integration is a practical and highly effective approach for modern assistive technologies.

XIII. FUTURE WORK

Wearable Integration: Deploy the system on smart glasses for portability and convenience.

- **Cloud Scalability:** Enable cloud-based processing for faster performance and broader deployment.
- **Source Code Input:** Extend functionality to analyze digital documents directly for accessibility.
- **Domain-Specific Models:** Fine-tune AI for healthcare, education, and public services to improve accuracy in specialized contexts.

REFERENCES

1. Catherine, ShiriAzenkot, Maya Cakmak, “Designing a Robot Guide for Blind People in Indoor Environments,” ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, 2015.

2. H. E. Chen, Y. Y. Lin, C. H. Chen, I. F. Wang, “Blindnavi: a mobile navigation app specially designed for the visually impaired,” ACM Conference Extended Abstracts on Human Factors in Computing Systems, 2015.

3. K. W. Chen, C. H. Wang, X. Wei, Q. Liang, C. S. Chen, M. H. Yang, and Y. P. Hung, “Vision based positioning for Internet-of-Vehicles,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, no.2, pp. 364–376, 2016.

4. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes Dataset for Semantic Urban Scene Understanding,” IEEE Conference on Computer Vision and Pattern Recognition, 2016.

5. J. Ducasse, M. Macé, M. Serrano, and C. Jouffrais, “Tangible Reels: Construction and Exploration of Tangible Maps by Visually Impaired Users,” ACM CHI Conference on Human Factors in Computing Systems, 2016.

6. J. Engel, T. Schops, and D. Cremers, “LSD-SLAM: Largescale direct monocular SLAM,” European Conference on Computer Vision, 2014.

7. S. Gilson, S. Gohil, F. Khan, V. Nagaonkar, “A Wireless Navigation System for the Visually Impaired,” Capstone Spring, 2015.

8. J. Guerreiro, D. Ahmetovic, K. M. Kitani, and C. Asakawa, “Virtual Navigation for Blind People: Building Sequential Representations of the Real-World,” International ACM SIGACCESS Conference on Computers and Accessibility, 2017.

9. Kendall, M. Grimes, and R. Cipolla, “PoseNet: a convolutional network for real-time 6-DOF camera relocalization,” International Conference on Computer Vision, 2015.

10. Kendall, and R. Cipolla, “Geometric loss function for camera pose regression with deep learning,” International Conference on Computer Vision, 2017.

11. R.Bharathi, T.Abirami,” Energy efficient compressive sensing with predictive model for IoT based medical data transmission”, Journal of Ambient Intelligence and Humanized Computing, November 2020, https://doi.org/10.1007/s12652-020-02670-z

12. R.Bharathi, T.Abirami,” Energy Efficient Clustering with Disease Diagnosis Model for IoT based Sustainable Healthcare Systems”, Sustainable Computing: Informatics and Systems, 23 September 2020, https://doi.org/10.1016/j.suscom.2020.100453